

## Indirect Inference for Stochastic Volatility Models via the Log-Squared Observations

By G. DHAENE\*



Geert Dhaene  
KULeuven, Departement Economi-  
sche Wetenschappen. Centrum voor  
Economische Studiën, Onderzoeks-  
groep Econometrie, Leuven.

### ABSTRACT

An indirect estimator of the stochastic volatility (SV) model with AR(1) log-volatility is proposed. The estimator is derived as an application of the method of indirect inference (Gouriéroux, Monfort and Renault (1993)), using an auxiliary SV model that mimics the SV model of interest (which has latent volatility) but is constructed so as to make volatility observable. The resulting estimator works by fitting an AR(1) to the log-squared observations and then applying a simple transformation to the parameter estimates. A closed-form expression for the asymptotic covariance matrix of the estimator is also derived. The estimator is applied to the Brussels All Shares Price Index from January 1, 1980, to January 16, 2003.

\* I thank the referee for helpful comments. Financial support from FWO research project G.0366.01 is gratefully acknowledged.

## I. INTRODUCTION

The phenomenon of volatility clustering is one of the most striking features of financial markets. While short-term returns on financial investment are typically uncorrelated over time and are found to be unpredictable, i.e. have a constant conditional mean given the past observations, there is overwhelming empirical evidence that the return variances are positively autocorrelated and predictable, i.e. the returns have a conditional variance that depends on past observations. Given the fundamental role that return variances and covariances play in portfolio management and asset pricing, it is important to understand their dynamic behaviour. At present, two classes of models have the inherent property of producing time-varying volatility, along with other phenomena often found in financial time series. The most popular of these is the class of (G)ARCH (Engle (1982); Bollerslev (1986)) and E-GARCH models (Nelson (1991)), which have the attractive feature of being easy to estimate. In these models, the return variance is driven by past shocks (essentially, the residuals) in the mean equation. By contrast, in SV models, which were introduced by Clark (1973) and extended by Tauchen and Pitts (1983), the return variance is modeled as a separate stochastic process, thus making the return variance a dynamic latent variable. As a result, SV models are much harder to estimate and have been used much less in applications. Following an important paper by Hull and White (1987), in which SV models appear as discrete time approximations to the continuous time volatility diffusions used in option pricing theory, there has been a renewed interest in SV models.

Considerable effort has been devoted to developing feasible techniques for estimating SV models. Taylor (1986) and Melino and Turnbull (1990) proposed GMM estimation based on the moments and autocovariances of the absolute returns. Jacquier, Polson and Rossi (1994), Andersen and Sørensen (1996, 1997) and Andersen, Chung and Sørensen (1999) used Monte Carlo methods to study the properties of these estimators. Other available estimation techniques for SV models include quasi-maximum likelihood (Nelson (1988); Harvey, Ruiz and Shephard (1994);

Ruiz (1994)), simulated maximum likelihood (Danielsson and Richard (1993); Danielsson (1994)), simulation-based GMM (Duffie and Singleton (1993)), indirect inference (Gouriéroux, Monfort and Renault (1993); Monfardini (1998)), Markov chain Monte Carlo methods (Jacquier, Polson and Rossi (1994); Kim, Shephard and Chib (1998); Chib, Nardari and Shephard (2002)), efficient method of moments (Gallant, Hsieh and Tauchen (1997); Andersen, Chung and Sørensen (1999)), ML Monte Carlo (Sandmann and Koopman (1998)) and (approximate) maximum likelihood (Fridman and Harris (1998)). With the exception of GMM and quasi-maximum likelihood, all of the existing methods require extensive numerical simulation and/or integration. Furthermore, obtaining accurate standard errors is far from simple, even with GMM (where the usual standard errors are found to be imprecise) or quasi-maximum likelihood (which involves the Kalman filter as an intermediary step in constructing the quasi-likelihood).

In this paper, a very simple estimator of the basic SV model is presented. In contrast with all existing estimators, closed-form expressions for the estimator and its asymptotic variance are obtained. The estimator is obtained by applying the method of indirect inference (Gouriéroux, Monfort and Renault (1993)) to an auxiliary SV model in which volatility is no longer latent, and then inverting the parameter estimates of the auxiliary model back to the parameters of the original SV model. The particular choice of auxiliary model allows all steps required in the indirect inference procedure to be carried out analytically.

The basic SV model is presented in Section II, along with its main characteristics. Section III briefly outlines the indirect inference approach and then applies it to the model at hand. In Section IV, the estimation method is illustrated with an application to the Brussels All Shares Price Index. Section V concludes. The more technical derivations are given in the Appendix.

## II. THE SV MODEL

In the basic SV model, the time series  $y_1, \dots, y_T$  is generated by

$$y_t = e^{h_t/2} u_t, \quad t = 1, \dots, T, \quad (1)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sqrt{\sigma^2(1 - \phi^2)} v_t, \quad (2)$$

$$h_1 \sim N(\mu, \sigma^2), \quad (3)$$

where  $u_t$  and  $v_t$  are standard normal variates, assumed to be mutually independent, independent of  $h_1$  and independent across time, where  $h_1, \dots, h_T$  is a latent (i.e. unobserved) time series, and where  $\phi$ ,  $\mu$  and  $\sigma^2$  are parameters.<sup>1</sup> In financial applications,  $y_t$  is typically the return in period  $t$  on a financial investment. The essential characteristic of the model is that the variance (i.e. the volatility) of  $y_t$  is governed by a separate stochastic process, which is given by (2)–(3). To see this more clearly, observe that the independence of  $u_t$  and  $v_{t-1}, v_{t-2}, \dots$  implies the independence of  $u_t$  and  $h_t$ . Therefore, the conditional mean and variance of  $y_t$ , given  $h_t$ , are

$$E[y_t | h_t] = e^{h_t/2} E[u_t] = 0 \quad (4)$$

and

$$\text{Var}[y_t | h_t] = E[y_t^2 | h_t] = e^{h_t} E[u_t^2] = e^{h_t} \quad (5)$$

for all  $t$ . Note also that  $y_t | h_t \sim N(0, e^{h_t})$ . Thus, the conditional mean of  $y_t$  is identically zero, and  $\log(\text{Var}[y_t | h_t]) = h_t$ , i.e.  $h_t$  is the log-volatility of  $y_t$ . The so-called mean equation (1) sets  $y_t$  equal to a standard normal variate  $u_t$  times the standard deviation  $e^{h_t/2}$ . Equation (2) specifies the log-volatility to be an AR(1) with autoregressive parameter  $\phi$ , unconditional mean  $\mu$  and unconditional variance  $\sigma^2$ . Equation (3) starts the autoregression of  $h_t$  by a draw from its stationary distribution. It is assumed that  $|\phi| < 1$ , thus ensuring that  $h_t$  (hence also  $y_t$ ) is

---

<sup>1</sup>It is more common to parameterise the model in terms of  $\phi$ ,  $\alpha = \mu(1 - \phi)$  and  $\omega = \sqrt{\sigma^2(1 - \phi^2)}$ . I prefer the parameterisation in terms of  $\phi$ ,  $\mu$  and  $\sigma^2$  for algebraic reasons and because of a parameter invariance result presented below.

stationary. The unconditional mean and variance of  $y_t$  are<sup>2</sup>

$$E[y_t] = 0$$

and

$$\text{Var}[y_t] = E[e^{h_t}] = e^{\mu + \frac{1}{2}\sigma^2}.$$

The latter equation follows from the well known property that  $h_t \sim N(\mu, \sigma^2)$  implies  $E[e^{h_t}]^r = E[e^{rh_t}] = e^{r\mu + \frac{1}{2}r^2\sigma^2}$  for any  $r$ . This property of the lognormal distribution will be used throughout the paper. The random variables  $u_t$  and  $v_t$  are sometimes called mean shocks and volatility shocks, respectively. The presence of a separate stochastic component  $v_t$  governing volatility (whence the name SV) constitutes the major difference of SV models relative to GARCH models. The latter class of models replace (2) by a specification in which  $h_{t+1}$  depends on  $u_t$  (and, possibly, on lags of  $h_t$  and  $u_t$ ) rather than on  $v_t$ . On the other hand, GARCH and SV models do share a number of important properties that are often found in financial time series data. First, there is no serial correlation in  $y_t$ , since

$$\text{Cov}[y_t, y_{t-p}] = E[y_t y_{t-p}] = E[e^{(h_t + h_{t-p})/2}] E[u_t] E[u_{t-p}] = 0$$

for any positive integer  $p$ . Secondly, there is serial correlation in  $y_t^2$ . To see this, note that  $\text{Cov}(h_t, h_{t-p}) = \phi^p \sigma^2$ , yielding  $h_t + h_{t-p} \sim N(2\mu, 2\sigma^2(1 + \phi^p))$ . So,

$$\begin{aligned} \text{Cov}[y_t^2, y_{t-p}^2] &= E[y_t^2 y_{t-p}^2] - E[y_t^2] E[y_{t-p}^2] \\ &= E[e^{h_t + h_{t-p}}] E[u_t^2] E[u_{t-p}^2] - e^{2\mu + \sigma^2} \\ &= e^{2\mu + \sigma^2(1 + \phi^p)} - e^{2\mu + \sigma^2}. \end{aligned}$$

For positive  $\phi$ ,  $\text{Cov}[y_t^2, y_{t-p}^2] > 0$  for any  $p$ . Positive serial correlation in  $y_t^2$ , coupled with the absence of serial correlation in  $y_t$ , is called volatility clustering, a phenomenon often observed

---

<sup>2</sup>A constant can be added to the right hand side of (1) if  $E[y_t] = 0$  is judged to be unrealistic. Equivalently, the time series  $y_t$  can first be demeaned.

in financial time series, where large returns of either sign tend to cluster together, as do small returns of either sign. Thirdly,

$$\frac{E[y_t^4]}{(\text{Var}[y_t])^2} = \frac{E[e^{2h_t}] E[u_t^4]}{e^{2\mu+\sigma^2}} = \frac{e^{2\mu+2\sigma^2} \cdot 3}{e^{2\mu+\sigma^2}} = 3e^{\sigma^2} > 3,$$

which shows that  $y_t$  has excess kurtosis.

From the point of view of inference, the fundamental problem with the SV model is the latent character of  $h_t$ , which makes it difficult to compute the values of the likelihood function and hence to estimate the parameters by maximum likelihood (ML). To see this, write the joint density of  $y_1, \dots, y_T$  and  $h_1, \dots, h_T$  as

$$\begin{aligned} f(y_1, \dots, y_T, h_1, \dots, h_T) &= f(h_1, \dots, h_T) f(y_1, \dots, y_T | h_1, \dots, h_T) \\ &= f(h_1) \left( \prod_{t=2}^T f(h_t | h_{t-1}) \right) \left( \prod_{t=1}^T f(y_t | h_t) \right). \end{aligned}$$

Now,

$$\begin{aligned} f(h_1) &= (2\pi\sigma^2)^{-1/2} e^{-(h_1-\mu)^2/(2\sigma^2)}, \\ f(h_t | h_{t-1}) &= [2\pi\sigma^2(1-\phi^2)]^{-1/2} e^{-[h_t-\mu-\phi(h_{t-1}-\mu)]^2/[2\sigma^2(1-\phi^2)]}, \\ f(y_t | h_t) &= (2\pi e^{h_t})^{-1/2} e^{-y_t^2/(2e^{h_t})}. \end{aligned}$$

So,

$$\begin{aligned} f(y_1, \dots, y_T, h_1, \dots, h_T) &= (2\pi)^{-T} (1-\phi^2)^{-(T-1)/2} e^{-\frac{1}{2} \sum_{t=1}^T h_t} \\ &\quad \times e^{-(h_1-\mu)^2/(2\sigma^2) - \sum_{t=2}^T [h_t-\mu-\phi(h_{t-1}-\mu)]^2/[2\sigma^2(1-\phi^2)]} \\ &\quad \times e^{-\sum_{t=1}^T y_t^2/(2e^{h_t})}. \end{aligned}$$

The likelihood function is the joint density of the observables  $y_1, \dots, y_T$  as a function of the parameters, i.e.

$$\begin{aligned} L(\phi, \mu, \sigma^2 | y_1, \dots, y_T) &= f(y_1, \dots, y_T) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(y_1, \dots, y_T, h_1, \dots, h_T) dh_1 \dots dh_T. \end{aligned}$$

Thus, the likelihood function involves a  $T$ -dimensional integral. This integral is not known to be expressible in terms of known mathematical functions. At present, numerical evaluation of the exact likelihood function is not feasible, because with the present speed of computers numerical integration is only possible over low-dimensional spaces, whereas in applications  $T$  is often large. In the next section, evaluation of the exact likelihood is avoided by recurrence to an auxiliary model which is easy to estimate, and whose parameter estimates can be transformed to yield estimates of  $\phi$ ,  $\mu$  and  $\sigma^2$ .

### III. INDIRECT INFERENCE

When the parameter vector  $\theta$  of a parametric model  $M$  is difficult to estimate by ML, indirect inference (Gouriéroux, Monfort and Renault (1993)) may be a feasible alternative to ML. The method involves the following steps:

- Estimate the parameter  $\theta_*$  of an auxiliary model  $M_A$ . Let  $\hat{\theta}_*$  be the estimate.
- Calculate the probability limit of  $\hat{\theta}_*$  under  $M$ , as a function of  $\theta$ . This gives  $\text{plim } \hat{\theta}_* = \theta_* = \theta_*(\theta)$ . For identification, it is assumed that  $D = \frac{\partial}{\partial \theta'} \theta_*(\theta)$  has full column rank.
- For a given non-stochastic positive definite weighting matrix  $W$ , solve

$$\min_{\theta} \left( \theta_*(\theta) - \hat{\theta}_* \right)' W \left( \theta_*(\theta) - \hat{\theta}_* \right).$$

The solution,  $\hat{\theta}$ , is the indirect estimator of  $\theta$ .

- Calculate the asymptotic covariance matrix  $\hat{\theta}$  as

$$V_{\theta} = (D'WD)^{-1}D'WV_{\theta_*}WD(D'WD)^{-1}, \quad (6)$$

where  $V_{\theta_*}$  is the asymptotic covariance matrix of  $\hat{\theta}_*$ .

Remarks:

- The function  $\theta_*(\cdot)$ , sometimes called the pseudo-true value function or binding function, links the parameter  $\theta_*$  to  $\theta$ . It is assumed that  $\theta_*(\cdot)$  exists, i.e. that it has a well-defined probability limit for all  $\theta$ . For identification,  $\theta_*(\cdot)$  must be injective, so it is required that  $\dim(\theta_*) \geq \dim(\theta)$ .
- The optimal weighting matrix, which gives the smallest  $V_\theta$ , is  $W = V_{\theta_*}^{-1}$ , in which case  $V_\theta = (D'V_{\theta_*}^{-1}D)^{-1}$ .
- When  $\dim(\theta_*) = \dim(\theta)$ ,  $\hat{\theta}$  does not depend on  $W$  and is equal to  $\theta_*^{-1}(\hat{\theta}_*)$ , the inverse of the pseudo-true value function at  $\hat{\theta}_*$ . In this case,  $V_\theta = D^{-1}V_{\theta_*}(D^{-1})'$ .
- The weighting matrix  $W$  and the pseudo-true value function  $\theta_*(\cdot)$  may be replaced by consistent estimates without affecting the asymptotic properties of  $\hat{\theta}$ .

Indirect inference will now be applied to estimate the parameter vector  $\theta = (\phi, \mu, \sigma^2)'$  of model  $M$ , which is defined by (1)–(3). As auxiliary model  $M_A$ , consider the SV model

$$y_t = e^{h_t/2} \epsilon_t, \quad (7)$$

$$h_{t+1} = \mu_* + \phi_*(h_t - \mu_*) + \sqrt{\sigma_*^2(1 - \phi_*^2)} v_t, \quad (8)$$

$$h_1 \sim N(\mu_*, \sigma_*^2), \quad (9)$$

where  $\epsilon_t$  and  $v_t$  are mutually independent, independent of  $h_1$  and independent across time,  $v_t$  is standard normal,  $\epsilon_t$  is a symmetric Bernoulli variate with  $\Pr[\epsilon_t = 1] = \Pr[\epsilon_t = -1] = 1/2$ , and  $\theta_* = (\phi_*, \mu_*, \sigma_*^2)'$  is the parameter vector. The only difference between this and the original SV model is that the normal variate  $u_t$  in (1) is replaced with a Bernoulli variate  $\epsilon_t$ . The key feature of the auxiliary model is that volatility now is observable, since (7) implies that  $h_t = \log y_t^2$ . Thus, ML estimation of  $\theta_*$  is straightforward. As an alternative to ML estimation, the autoregression

$$\log y_t^2 = \mu_* + \phi_*(\log y_{t-1}^2 - \mu_*) + \sqrt{\sigma_*^2(1 - \phi_*^2)} v_{t-1}, \quad t = 2, \dots, T \quad (10)$$



can be fitted by (non-linear) least squares. Let  $\hat{\theta}_* = (\hat{\phi}_*, \hat{\mu}_*, \hat{\sigma}_*^2)'$  be the resulting estimator.<sup>3</sup> The ML and the non-linear least squares estimators are asymptotically equivalent in this case.<sup>4</sup> As  $T$  grows large,  $\hat{\phi}_*$ ,  $\hat{\mu}_*$  and  $\hat{\sigma}_*^2$  converge to their population counterparts (or probability limits). For an autoregression like (10), the population counterparts are straightforward. The estimator  $\hat{\phi}_*$  is the sample first-order autocorrelation of  $\log y_t^2$  and hence converges to the population first-order autocorrelation. That is,

$$\text{plim } \hat{\phi}_* = \phi_* = \frac{\gamma_1}{\gamma_0}, \quad (11)$$

where  $\gamma_j = \text{Cov}(\log y_t^2, \log y_{t-j}^2)$ . It is important to note that, in deriving (11), it was not assumed that (10) is correctly specified. Indeed, from the viewpoint of  $M$ , (10) *is misspecified*, but still we know that  $\hat{\phi}_*$ , being a function of sample moments, converges to the same function of the corresponding population moments. Furthermore, note that the same notation, i.e.  $\phi_*$ , has been used for a parameter in a misspecified model and for the probability limit of its estimator. The latter is called the pseudo-true value, to emphasise the fact that the model is - or may be - misspecified. By similar reasoning,  $\hat{\mu}_*$  is asymptotically equivalent to the sample mean of  $\log y_t^2$  and hence converges to

$$\text{plim } \hat{\mu}_* = \mu_* = m, \quad (12)$$

where  $m = E(\log y_t^2)$ , the unconditional population mean of  $\log y_t^2$ . To find the probability limit of  $\hat{\sigma}_*^2$  in terms of population moments, note that the residual variance of (10) is  $\hat{\sigma}_*^2(1 - \hat{\phi}_*^2)$ , since  $v_{t-1}$  has unit variance by assumption. This residual variance is the sample variance of  $\log y_t^2 - \hat{\mu}_* - \hat{\phi}_*(\log y_{t-1}^2 - \hat{\mu}_*)$  and so

---

<sup>3</sup>Note that the non-linear least squares estimates  $\hat{\phi}_*$ ,  $\hat{\mu}_*$  and  $\hat{\sigma}_*^2$  can also be obtained as  $\hat{\phi}_*$ ,  $\hat{\alpha}_*(1 - \hat{\phi}_*)^{-1}$  and  $\hat{\omega}_*^2(1 - \hat{\phi}_*^2)^{-1}$ , where  $\hat{\alpha}_*$  and  $\hat{\phi}_*$  are the (linear) least squares estimates in  $\log y_t^2 = \hat{\alpha}_* + \hat{\phi}_* \log y_{t-1}^2 + \text{residual}$ , and  $\hat{\omega}_*^2$  is the average of the squared residuals.

<sup>4</sup>The only difference between the two methods is the treatment of the first observation ( $t = 1$ ). The non-linear least-squares estimator “looses” the first observation (although this can be avoided), while ML exploits the fact that  $h_1 \sim N(\mu_*, \sigma_*^2)$ . This difference is negligible as  $T$  becomes large.

converges to  $\text{Var}(\log y_t^2 - \mu_* - \phi_*(\log y_{t-1}^2 - \mu_*)) = \text{Var}(\log y_t^2)(1 - \phi_*^2) = \gamma_0(1 - \phi_*^2)$ . Therefore,

$$\text{plim } \hat{\sigma}_*^2 = \sigma_*^2 = \gamma_0. \quad (13)$$

The parameters  $\phi_*$ ,  $\mu_*$  and  $\sigma_*^2$  are now to be expressed in terms of  $\phi$ ,  $\mu$  and  $\sigma^2$ , the parameters of  $M$ . In view of (1),  $M$  implies that  $\log y_t^2 = h_t + \log u_t^2$  and hence

$$m = \mu + c_1, \quad \gamma_0 = \sigma^2 + c_2, \quad \gamma_1 = \phi\sigma^2, \quad (14)$$

where, as shown in the Appendix,

$$c_1 = E(\log u_t^2) = -1.270$$

and

$$c_2 = \text{Var}(\log u_t^2) = 4.935.$$

Hence, the components of the pseudo-true value function  $\theta_*(\theta)$  are

$$\phi_* = \phi \frac{\sigma^2}{\sigma^2 + c_2}, \quad \mu_* = \mu + c_1, \quad \sigma_*^2 = \sigma^2 + c_2. \quad (15)$$

Solving for  $\phi$ ,  $\mu$  and  $\sigma^2$  (i.e. inverting  $\theta_*(\cdot)$ ) gives

$$\phi = \phi_* \frac{\sigma_*^2}{\sigma_*^2 - c_2}, \quad \mu = \mu_* - c_1, \quad \sigma^2 = \sigma_*^2 - c_2. \quad (16)$$

Substituting  $\hat{\phi}_*$ ,  $\hat{\mu}_*$  and  $\hat{\sigma}_*^2$  for  $\phi_*$ ,  $\mu_*$  and  $\sigma_*^2$  on the right-hand sides of (16) yields  $\hat{\theta} = \theta_*^{-1}(\hat{\theta}_*)$ . This gives the indirect estimators  $\hat{\phi}$ ,  $\hat{\mu}$  and  $\hat{\sigma}^2$ , which consistently estimate  $\phi$ ,  $\mu$  and  $\sigma^2$ . No weighting matrix is needed here, because  $\dim(\theta_*) = 3 = \dim(\theta)$ .

It is worth noting that, while evaluation of the likelihood function of  $M$  at different values of  $\theta$  is not feasible, the pseudo-true value function, which links  $\theta_*$  to  $\theta$ , can be calculated analytically and is remarkably simple. Furthermore, the estimator  $\hat{\theta}$ , which is derived here as an indirect estimator, can also be viewed as an application of Gallant and Tauchen's (1996) method for generating moment conditions from the score function of an auxiliary model  $M_A$ . These moment conditions turn out to be simple to

handle analytically under the structural model  $M$ , while the likelihood function of  $M$  is not tractable. To see this, consider the contribution of observation  $t \geq 2$  to the score function of  $M_A$ :

$$\begin{pmatrix} \phi_*(1 - \phi_*^2)^{-1} - \sigma_*^{-2}(1 - \phi_*^2)^{-2}(d_t - \phi_*d_{t-1})(\phi_*d_t - d_{t-1}) \\ \sigma_*^{-2}(1 + \phi_*)^{-1}(d_t - \phi_*d_{t-1}) \\ -\frac{1}{2}\sigma_*^{-2} + \frac{1}{2}\sigma_*^{-4}(1 - \phi_*^2)^{-1}(d_t - \phi_*d_{t-1})^2 \end{pmatrix},$$

where  $d_t = \log y_t^2 - \mu_*$ . Taking expectations under  $M$ , equating to zero and solving for  $\mu$ ,  $\phi$  and  $\sigma^2$  gives (16). A third, and obvious, interpretation of  $\hat{\theta}$  is as a method of moments estimator. Considering that (14) establishes a direct link between  $\theta$  and the population moments  $\eta = (m, \gamma_0, \gamma_1)'$ , one can directly solve (14) to yield

$$\phi = \frac{\gamma_1}{\gamma_0 - c_2}, \quad \mu = m - c_1, \quad \sigma^2 = \gamma_0 - c_2. \quad (17)$$

These expressions are equivalent to (16). Substituting sample moments  $\hat{\eta} = (\hat{m}, \hat{\gamma}_0, \hat{\gamma}_1)'$  for population moments on the right-hand sides of (17) yields a method of moments estimator of  $\theta$  that is asymptotically equivalent to  $\hat{\theta}$ .

The asymptotic covariance matrix  $V_\theta$  of  $\hat{\theta}$  is obtained by applying (18) after calculating the asymptotic covariance matrix  $V_{\theta_*}$  of  $\hat{\theta}_*$ . The latter matrix is found by calculating the asymptotic covariance matrix  $V_\eta$  of  $\hat{\eta}$  and applying the delta method to the transformation  $\eta \rightarrow \theta_*$ .<sup>5 6</sup> Let  $V_\eta(i, j)$  be the  $(i, j)$ -th

---

<sup>5</sup>An asymptotic covariance matrix, say  $V_\theta$ , is defined as  $\lim_{T \rightarrow \infty} \text{Var}[\sqrt{T}(\hat{\theta} - \theta)]$ .

<sup>6</sup>The intermediary step of deriving  $V_{\theta_*}$  is not needed in this particular case, since one can apply the delta method directly to the transformation  $\eta \rightarrow \theta$ , which is given by (17). For the sake of illustrating the logic of indirect inference, however,  $V_{\theta_*}$  will be derived.

element of  $V_\eta$ . It is shown in the Appendix that

$$\begin{aligned}
V_\eta(1, 1) &= \frac{1 + \phi}{1 - \phi} \sigma^2 + c_2, \\
V_\eta(2, 2) &= 2 \frac{1 + \phi^2}{1 - \phi^2} \sigma^4 + 4c_2 \sigma^2 + c_4 - c_2^2, \\
V_\eta(3, 3) &= \left( 1 + \phi^2 + 4 \frac{\phi^2}{1 - \phi^2} \right) \sigma^4 + 2c_2(1 + \phi^2) \sigma^2 + c_2^2, \\
V_\eta(2, 1) &= c_3, \\
V_\eta(3, 1) &= 0, \\
V_\eta(3, 2) &= 2\phi \left( 1 + \frac{1 + \phi^2}{1 - \phi^2} \right) \sigma^4 + 4c_2 \phi \sigma^2,
\end{aligned}$$

where

$$c_3 = E(\log u_t^2 - c_1)^3 = -16.83$$

and

$$c_4 = E(\log u_t^2 - c_1)^4 = 170.5.$$

In view of (11)–(14), the transformation  $\eta \rightarrow \theta_*$  has the Jacobian matrix

$$\begin{aligned}
J &= \frac{\partial \theta_*}{\partial \eta'} = \begin{pmatrix} 0 & -\gamma_1 \gamma_0^{-2} & \gamma_0^{-1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & -\phi \sigma^2 (\sigma^2 + c_2)^{-2} & (\sigma^2 + c_2)^{-1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}
\end{aligned}$$

and so  $V_{\theta_*}$  is obtained as  $JV_\eta J'$ . Finally, from (15),

$$D = \frac{\partial}{\partial \theta'} \theta_*(\theta) = \begin{pmatrix} \frac{\sigma^2}{\sigma^2 + c_2} & 0 & \phi c_2 (\sigma^2 + c_2)^{-2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the lower triangular elements of  $V_\theta = D^{-1}JV_\eta J'(D^{-1})'$  are

$$\begin{pmatrix} \frac{(1-\phi^2)(\sigma^2+c_2)^2+\phi^2c_4}{\sigma^4} & \cdot & \cdot \\ -\frac{\phi}{\sigma^2}c_3 & \frac{1+\phi}{1-\phi}\sigma^2+c_2 & \cdot \\ 2\phi\sigma^2-\frac{\phi}{\sigma^2}(c_4-c_2^2) & c_3 & 2\frac{1+\phi^2}{1-\phi^2}\sigma^4+4\sigma^2c_2+c_4-c_2^2 \end{pmatrix}. \tag{18}$$

It is of interest to note that  $V_\theta$  does not depend on  $\mu$ , and that the asymptotic variances of  $\hat{\mu}$  and  $\hat{\sigma}^2$  are unbounded as  $\phi \rightarrow 1$ . The asymptotic variance of  $\hat{\phi}$ , however, remains bounded as  $\phi \rightarrow 1$ .

#### IV. APPLICATION TO THE BRUSSELS ALL SHARES PRICE INDEX

Let  $r_t$  be the return on the Belgian All Shares Price Index between successive trading days  $t-1$  and  $t$ , with  $t-1$  ranging from December 31, 1979, to January 15, 2003. The data were taken from Datastream with zero returns removed, as these correspond to non-trading weekdays or, almost certainly, to errors. This yielded a total of  $T = 5627$  non-zero returns, and an average yearly (ex-dividend) return equal to  $\frac{1}{23.04} \sum_t r_t = 0.0737$ . Let  $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$ . Descriptive statistics on the daily returns  $r_t$ , the squared de-meaned returns  $(r_t - \bar{r})^2$  and the log-squared de-meaned returns  $\log(r_t - \bar{r})^2$  are given in Table 1.

Table 1: Descriptive statistics

	$r_t$	$(r_t - \bar{r})^2$	$\log(r_t - \bar{r})^2$
mean	$3.02 \times 10^{-4}$	$7.66 \times 10^{-6}$	-11.45
std. deviation	$8.75 \times 10^{-3}$	$2.89 \times 10^{-4}$	2.498
skewness	-0.342	19.80	-1.109
kurtosis	12.20	667.2	2.522

Fitting (10) with  $y_t = r_t - \frac{1}{T} \sum_{t=1}^T r_t$  gives

$$\begin{aligned} \log y_t^2 &= -11.45 + 0.1959(\log y_{t-1}^2 + 11.45) \\ &\quad + \sqrt{6.239(1 - (0.1959)^2)} \hat{v}_{t-1}, \end{aligned}$$

for  $t = 2, \dots, T$ , where, by construction,  $\frac{1}{T-1} \sum_{t=2}^T \hat{v}_{t-1}^2 = 1$ . The indirect estimates of  $\phi$ ,  $\mu$  and  $\sigma^2$  and their standard errors now follow from (16) and (18)<sup>7</sup>:

$$\begin{aligned} \hat{\phi} &= 0.1959 \frac{6.239}{6.239 - 4.935} = 0.937, & \text{st.err.}(\hat{\phi}) &= 0.127, \\ \hat{\mu} &= -11.45 + 1.270 = -10.18, & \text{st.err.}(\hat{\mu}) &= 0.090, \\ \hat{\sigma}^2 &= 6.239 - 4.935 = 1.304, & \text{st.err.}(\hat{\sigma}^2) &= 0.194. \end{aligned}$$

The estimate of  $\phi$ , which is close to but smaller than 1, is in line with estimates that have been reported in the literature. The relatively large standard errors of the estimates result from the fact that  $\phi$  appears to be close to 1, and from the fact that the indirect estimator exploits only the information contained in the mean, variance and first-order autocorrelation of  $\log y_t^2$ .

## V. CONCLUSION

Although SV models are notoriously difficult to estimate, the use of a judiciously chosen auxiliary model and the application of the method of indirect inference yields an estimator and an associated asymptotic covariance matrix that have simple closed-form expressions. Unfortunately, this comes at a price: the resulting estimator is very inefficient. A preliminary comparison with Monte Carlo results by Jacquier, Polson and Rossi (1994) shows that the standard errors of the estimators presented here may be up to 100 times as large as those of the Markov Chain Monte

---

<sup>7</sup>Standard errors are computed as the square-root of  $T - 1$  times the appropriate element of  $V_\theta$ , with estimates replacing parameters.

Carlo estimator (though under unfavourable conditions). Therefore, the exploitation of the additional information contained in higher-order autocorrelations of  $\log y_t^2$ , or in moments and autocorrelations of  $|y_t|$ , will reduce the variance of the estimator considerably. It is possible to obtain closed-form expressions both for the optimal weighting matrix of the GMM estimator in this context and for its asymptotic covariance matrix. I hope to report on this in the near future.

It would be natural to test, for example, whether the AR(1) specification for the log-volatility is not too restrictive, or whether the normality assumption of  $u_t$  in equation (1) is realistic. While likelihood-based testing methods are presently not feasible, GMM-based methods are relatively straightforward. It is not difficult to derive moment conditions for an AR(2) specification for log-volatility, hence the standard GMM estimates and standard errors yield a test of the AR(1) specification. Furthermore, assuming that the log-volatility is correctly specified, that  $h_t$  and  $u_t$  are independent processes and that  $u_t$  is i.i.d., the moment conditions derived from the expectation and autocorrelations of  $y_t^2$  are solely based on the second moment of  $u_t$  (which equals one, without loss of generality). Adding moments conditions derived from the expectation and autocorrelations of other powers of  $|y_t|$ , the GMM test for overidentifying restrictions is a test of the normality of  $u_t$ .

## APPENDIX

### Calculation of $c_1, \dots, c_4$

For any positive integer  $n$ , upon substituting  $t = x^2/2$ ,

$$\begin{aligned} \int_{-\infty}^{\infty} \left( \log \frac{x^2}{2} \right)^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx &= 2 \int_0^{\infty} \left( \log \frac{x^2}{2} \right)^n \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{\pi}} \int_0^{\infty} (\log t)^n t^{-1/2} e^{-t} dt \\ &= \frac{\Gamma^{(n)}\left(\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}, \end{aligned}$$

where  $\Gamma^{(n)}(z)$  is the  $n$ -th derivative of  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ , the gamma function, and  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ . See Abramowitz and Stegun (1970) for properties and values of the gamma and related functions. Now,

$$\begin{aligned} c_1 &= \int_{-\infty}^{\infty} (\log x^2) \left( \frac{1}{\sqrt{2\pi}} \right) e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \left( \log \frac{x^2}{2} \right) \left( \frac{1}{\sqrt{2\pi}} \right) e^{-x^2/2} dx + \log 2 \\ &= \psi\left(\frac{1}{2}\right) + \log 2, \end{aligned}$$

where  $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ , the digamma function. For  $n = 2, 3, 4$ ,

$$\begin{aligned} c_n &= \int_{-\infty}^{\infty} (\log x^2 - c_1)^n \left( \frac{1}{\sqrt{2\pi}} \right) e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \left( \log \frac{x^2}{2} - \psi\left(\frac{1}{2}\right) \right)^n \left( \frac{1}{\sqrt{2\pi}} \right) e^{-x^2/2} dx \\ &= g_n\left(\frac{1}{2}\right), \end{aligned}$$

where

$$g_n(z) = \sum_{i=0}^n \binom{n}{i} \frac{\Gamma^{(i)}(z)}{\Gamma(z)} (-\psi(z))^{n-i}.$$

Some tedious but straightforward algebra shows that

$$\begin{aligned} g_2(z) &= \psi'(z), \\ g_3(z) &= \psi''(z), \\ g_4(z) &= \psi'''(z) + 3(\psi'(z))^2, \end{aligned}$$

with primes denoting derivatives. Now,  $\psi(\frac{1}{2}) = -\gamma - 2\log 2$ , where  $\gamma = 0.5772$  is Euler's constant,  $\psi'(\frac{1}{2}) = \frac{\pi^2}{2}$ ,  $\psi''(\frac{1}{2}) = -14\zeta(3)$ , where  $\zeta(3) = 1.202$  is the value of the Riemann zeta function at 3, and  $\psi'''(\frac{1}{2}) = \pi^4$ . It follows that

$$\begin{aligned} c_1 &= -\log 2 - \gamma = -1.270, \\ c_2 &= \frac{1}{2}\pi^2 = 4.935, \\ c_3 &= -14\zeta(3) = -16.83, \end{aligned}$$



and

$$c_4 = \frac{7}{4}\pi^4 = 170.5.$$

### Calculation of $V_\eta$

The elements of  $\hat{\eta}$  are the sample mean, variance and first-order autocovariance of  $\log y_t^2$ . Thus, the asymptotic covariance matrix is

$$V_\eta = \lim_{T \rightarrow \infty} \text{Var}[\sqrt{T}(\hat{\eta} - \eta)] = \sum_{j=-\infty}^{\infty} \text{Cov}(\xi_t, \xi_{t-j}),$$

where

$$\xi_t = \begin{pmatrix} z_t \\ z_t^2 \\ z_t z_{t-1} \end{pmatrix}$$

and

$$z_t = \log y_t^2 - \mu - c_1.$$

Write  $z_t$  as  $k_t + w_t$ , where  $k_t = h_t - \mu$  and  $w_t = \log u_t^2 - c_1$ . Now,  $w_t$  and  $k_t$  have zero mean and are independent, and we have that  $\text{Cov}(k_t, k_{t-i}) = \phi^{|i|}\sigma^2$ ,  $\text{Cov}(k_t k_{t-i}, k_{t-j}) = 0$ , and  $\text{Cov}(k_t k_{t-i}, k_{t-j} k_{t-l}) = (\phi^{|j|+|i-l|} + \phi^{|l|+|i-j|})\sigma^4$  for any integers  $i, j$  and  $l$ . Using these properties, the elements of  $V_\eta$  are found as

$$\begin{aligned} V_\eta(1, 1) &= \sum_{j=-\infty}^{\infty} \text{Cov}(k_t + w_t, k_{t-j} + w_{t-j}) \\ &= \sum_{j=-\infty}^{\infty} \{\text{Cov}(k_t, k_{t-j}) + \text{Cov}(w_t, w_{t-j})\} = \frac{1+\phi}{1-\phi}\sigma^2 + c_2, \end{aligned}$$

$$\begin{aligned}
V_\eta(2, 2) &= \sum_{j=-\infty}^{\infty} \text{Cov} [(k_t + w_t)^2, (k_{t-j} + w_{t-j})^2] \\
&= \sum_{j=-\infty}^{\infty} \{ \text{Cov}(k_t^2, k_{t-j}^2) + 4 \text{Cov}(k_t w_t, k_{t-j} w_{t-j}) \\
&\quad + \text{Cov}(w_t^2, w_{t-j}^2) \} \\
&= 2 \frac{1 + \phi^2}{1 - \phi^2} \sigma^4 + 4c_2 \sigma^2 + c_4 - c_2^2, \\
V_\eta(3, 3) &= \sum_{j=-\infty}^{\infty} \text{Cov} [(k_t + w_t)(k_{t-1} + w_{t-1}), \\
&\quad (k_{t-j} + w_{t-j})(k_{t-j-1} + w_{t-j-1})] \\
&= \sum_{j=-\infty}^{\infty} \{ \text{Cov}(k_t k_{t-1}, k_{t-j} k_{t-j-1}) \\
&\quad + \text{Cov}(k_t w_{t-1}, k_{t-j} w_{t-j-1}) \\
&\quad + \text{Cov}(k_t w_{t-1}, w_{t-j} k_{t-j-1}) \\
&\quad + \text{Cov}(w_t k_{t-1}, k_{t-j} w_{t-j-1}) \\
&\quad + \text{Cov}(w_t k_{t-1}, w_{t-j} k_{t-j-1}) \\
&\quad + \text{Cov}(w_t w_{t-1}, w_{t-j} w_{t-j-1}) \} \\
&= \left( 1 + \phi^2 + 4 \frac{\phi^2}{1 - \phi^2} \right) \sigma^4 + 2c_2(1 + \phi^2) \sigma^2 + c_2^2, \\
V_\eta(2, 1) &= \sum_{j=-\infty}^{\infty} \text{Cov} [(k_t + w_t)^2, k_{t-j} + w_{t-j}] \\
&= \sum_{j=-\infty}^{\infty} \text{Cov}(w_t^2, w_{t-j}) = c_3, \\
V_\eta(3, 1) &= \sum_{j=-\infty}^{\infty} \text{Cov} [(k_t + w_t)(k_{t-1} + w_{t-1}), k_{t-j} + w_{t-j}] \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
V_\eta(3, 2) &= \sum_{j=-\infty}^{\infty} \text{Cov} [(k_t + w_t)(k_{t-1} + w_{t-1}), (k_{t-j} + w_{t-j})^2] \\
&= \sum_{j=-\infty}^{\infty} \{ \text{Cov}(k_t k_{t-1}, k_{t-j}^2) + 2 \text{Cov}(k_t w_{t-1}, k_{t-j} w_{t-j}) \\
&\quad + 2 \text{Cov}(k_{t-1} w_t, k_{t-j} w_{t-j}) \} \\
&= 2\phi \left( 1 + \frac{1 + \phi^2}{1 - \phi^2} \right) \sigma^4 + 4c_2 \phi \sigma^2.
\end{aligned}$$

## REFERENCES

- Abramowitz, M. and I.A. Stegun, 1970, *Handbook of Mathematical Functions*, (Dover Publications, New York).
- Andersen, T.G., H.-J. Chung and B.E. Sørensen, 1999, Efficient Method of Moments Estimation of a Stochastic Volatility Model: a Monte Carlo Study, *Journal of Econometrics* 91, 61-87.
- Andersen, T.G. and B.E. Sørensen, 1996, GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study, *Journal of Business & Economic Statistics* 14, 328-352.
- Andersen, T.G. and B.E. Sørensen, 1997, GMM and QML Asymptotic Standard Deviations in Stochastic Volatility Models: Comments on Ruiz (1994), *Journal of Econometrics* 76, 397-403.
- Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* 31, 307-327.
- Chib, S., F. Nardari and N. Shephard, 2002, Markov Chain Monte Carlo Methods for Stochastic Volatility Models, *Journal of Econometrics* 108, 281-316.
- Clark, P.K., 1973, A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica* 41, 135-155.
- Danielsson, J., 1994, Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood, *Journal of Econometrics* 54, 375-400.
- Danielsson, J. and J.F. Richard, 1993, Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models, *Journal of Applied Econometrics* 8, Suppl., S153-S173.
- Duffie, D. and K.J. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929-952.
- Engle, R.F., 1982, Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United-Kingdom Inflation, *Econometrica* 50, 987-1007.
- Fridman, M. and L. Harris, 1998, A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models, *Journal of Business & Economic Statistics* 16, 284-291.
- Gallant, A.R., D.A. Hsieh and G.E. Tauchen, 1997, Estimation of Stochastic Volatility Models with Diagnostics, *Journal of Econometrics* 81, 159-192.
- Gallant, A.R. and G.E. Tauchen, 1996, Which Moments to Match? *Econometric Theory* 12, 657-681.
- Gouriéroux, C., A. Monfort and E. Renault, 1993, Indirect Inference, *Journal of Applied Econometrics* 8, S85-S118.

- Harvey, A.C., E. Ruiz and N. Shephard, 1994, Multivariate Stochastic Volatility Models, *Review of Economic Studies* 61, 247-264.
- Hull, J. and A. White, 1987, The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance* 17, 281-300.
- Jacquier, E., N.G. Polson and P.E. Rossi, 1994, Bayesian Analysis of Stochastic Volatility Models, *Journal of Business & Economic Statistics* 12, 69-87.
- Kim, S., N.G. Shephard and S. Chib, 1998, Stochastic Volatility: Optimal Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies* 45, 361-393.
- Melino, A. and S.M. Turnbull, 1990, Pricing Foreign Currency Options with Stochastic Volatility, *Journal of Econometrics* 45, 239-265.
- Monfardini, C., 1998, Estimating Stochastic Volatility Models through Indirect Inference, *Econometrics Journal* 1, C113-C128.
- Nelson, D.B., 1988, Time Series Behavior of Stock Market Volatility and Returns, Ph.D. dissertation, (MIT).
- Nelson, D.B., 1991, Conditional Heteroskedasticity in Asset Returns: a New Approach, *Econometrica* 59, 347-370.
- Ruiz, E., 1994, Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models, *Journal of Econometrics* 63, 289-306.
- Sandmann, G. and S.J. Koopman, 1998, Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, *Journal of Econometrics* 87, 271-301.
- Tauchen, G.E. and M. Pitts, 1983, The Price Variability-Volume Relationship on Speculative Markets, *Econometrica* 51, 484-505.
- Taylor, S., 1986, Modelling Financial Time Series, (Wiley, Chichester).